

Esquemas de comunicação no processamento de linguagem natural

Communication schemes in natural language processing

<https://doi.org/10.34112/2317-0972a2022v40n86p41-54>

PAULO HENRIQUE SOUTO MAIOR SERRANO¹

RESUMO: Partindo da apresentação do modelo matemático da comunicação, das funções de linguagem e do processo semiótico de comunicação, o artigo considera a compreensão do ruído como parte intrínseca do processo comunicativo. A partir disso, procura discutir a forma como a linguagem tem sido analisada pelas máquinas. Considerações sobre o processamento de linguagem natural são realizadas a partir da apresentação do *software* livre denominado dadossemiotica. A análise do discurso pela linguística computacional e a partir da teoria semiótica francesa constitui uma abordagem relevante para lidar com os ruídos inerentes ao processo de comunicação.

PALAVRAS-CHAVE: dadossemiotica; processamento de linguagem natural; semiótica greimasiana; esquema de comunicação.

ABSTRACT: The article presents the mathematical model of communication, language functions and the semiotic communication process, presenting the understanding of noise as an intrinsic part of the process and considering, by this understanding, the way in which language has been analyzed by machines. Considerations about natural language processing are conducted from the presentation of the open source software called dadossemiotica. Discourse analysis from computational linguistics and from french

1. Universidade Federal da Paraíba, João Pessoa-PB, Brasil.

semiotic theory constitutes a relevant approach to deal with the noise inherent in the communication process.

KEYWORDS: dadossemiotica; natural language processing; greimasian semiotics; communication scheme.

INTRODUÇÃO

As teorias e esquemas que buscaram explicar a comunicação e a linguagem compõem um conjunto de conhecimentos relevantes para a compreensão das tecnologias de processamento de linguagem natural. O processamento computacional da linguagem encontra no ruído um enorme condicionante para a sua eficácia, de modo semelhante aos diálogos humanos compreendidos pelos modelos que apresentam a comunicação como um processo constantemente sujeito à presença de ruídos.

Este artigo busca compreender como os esquemas de comunicação podem esclarecer as dificuldades e limitações intrínsecas ao processo comunicacional, através de sua aplicação no processamento de linguagem natural em soluções tecnológicas que buscam analisar o discurso ou responder às intenções de usuários que interagem com robôs de conversação. A correspondência das teorias que explicam o processo comunicacional com iniciativas computacionais será apresentada a partir do *software* livre dadossemiotica, desenvolvido pelo grupo Texto Livre.

A fundamentação teórica apresenta um panorama histórico importante para a compreensão do processo comunicacional; ela parte da teoria matemática da comunicação de Claude Shannon e Warren Weaver (1964), publicada originalmente em 1948 e focada no volume de informação do canal de comunicação e no ruído que poderia fazer parte desse processo.

Na sequência do artigo são apresentados a contribuição de Roman Jakobson (1960) e o modelo de Ignácio Assis Silva segundo Matte (2008). A contribuição do modelo brasileiro relaciona-se com a inclusão do ruído semiótico, um tipo de ruído presente no que não é expresso no processo comunicacional. Essa contribuição teórica pode ser relevante para um dos principais desafios da linguística computacional moderna, que é o de lidar com os pressupostos que a máquina deve interpretar.

O *software* livre dadossemiotica é uma solução educacional, voltada para a análise de textos, que se utiliza de uma abordagem estatística para a análise semântica e da quebra de frases em unidades menores, chamada tokenização, para a análise sintática do texto.

Essa solução idealizada pelo grupo de pesquisa Texto Livre está sendo desenvolvida para a implementação no sistema de gerenciamento de aprendizagem Moodle.

A semiótica francesa e os procedimentos para a análise de discurso através da linguística computacional são considerados uma abordagem relevante para lidar com os ruídos inerentes ao processo de comunicação, sobretudo aquele provocado pelas lacunas deixadas pelos interlocutores.

Destinador e destinatário dependem de informações contextuais para se comunicar adequadamente. Por exemplo, quando um lembrete, com horário específico para as dez horas do dia seguinte, é solicitado por voz para alguma assistente virtual, a omissão do turno do horário na composição do pedido é solucionada pelo destinatário quando este pergunta se é dez da manhã ou da noite – um processo de desambiguação que reduz a eficiência da comunicação.

Compreender as lacunas deixadas pelo destinador ou lidar com o imprevisível são alguns dos desafios que a linguística computacional precisa solucionar para oferecer serviços cada vez mais completos e satisfatórios. Espera-se que as contribuições teóricas deste artigo sobre as funções de linguagem, os diferentes tipos de ruído – dentre eles o ruído semiótico, trazido à tona por meio da semiótica greimasiana – e o desenvolvimento do *software* dadossemiótica possam oferecer caminhos para uma análise robusta e escalável mais eficiente.

TEORIA MATEMÁTICA DA COMUNICAÇÃO

A importância e valorização que os matemáticos e físicos receberam após a Segunda Guerra Mundial e durante a Guerra Fria foram condicionantes para o desenvolvimento de teorias e tecnologias relacionadas à telecomunicação, cibernética, computação, estatística, entre outras. A criptografia e a descriptografia se tornaram competências extremamente desejáveis em um contexto de tensões e espionagens entre países.

Integrando esse conjunto de matemáticos estavam Claude Shannon e Warren Weaver, que procuraram pensar o processo de comunicação a partir de uma lógica matemática e direcionaram seus estudos para o canal de comunicação. Quantidade de informação, capacidade, codificação, qualidade de transmissão de sinal e ruído – elementos estes ligados à linguagem – foram questões respondidas pelos pesquisadores em sua obra *A Teoria Matemática da Comunicação*, publicada em 1948.

Claude Shannon discute conceitos e cálculos matemáticos que estabelecem as bases de uma teoria da informação, procurando entender como fluxos

comunicacionais e computacionais de informação contínua poderiam ser otimizados para reduzir o ruído e a entropia da informação ou a incerteza com que um símbolo ou unidade de informação é esperada. Por exemplo, em uma comunicação comum em português são esperadas vogais depois de consoantes; identificar a probabilidade do surgimento dessas sequências de símbolos reduziria a entropia da comunicação, tornando-a mais eficiente (SHANNON; WEAVER, 1964, p. 53).

A teoria de Shannon se apresenta em um paradigma científico positivista, com abstrações matemáticas que descrevem e exemplificam sua proposta e capacidade de mensuração do processo comunicacional. Warren Weaver, por outro lado, oferece na obra uma perspectiva mais ampla e concreta para uma ampliação da compreensão das ideias e propostas mais técnicas de Shannon. A sistematização e organização trazidas por Weaver contribuíram para a popularização da teoria matemática da comunicação e consolidação da comunicação enquanto ciência. O modelo ou esquema de comunicação proposto pelos autores, apresentado na imagem 1, tornou-se muito popular, influenciando, de certo modo, a forma de exibição de modelos posteriores.

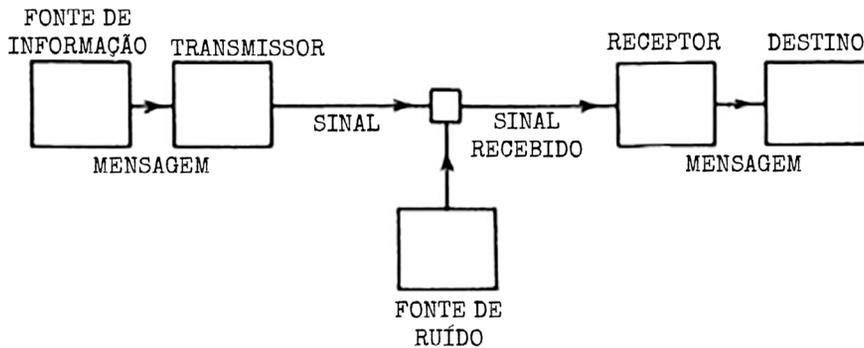


Imagem 1: Esquema de comunicação de Shannon e Weaver
Fonte: Shannon; Weaver (1964, p. 7) com tradução do autor.

De acordo com o modelo, a fonte de informação seleciona uma mensagem, em um universo de mensagens possíveis, o transmissor transforma a mensagem em um sinal e a envia por meio de um canal de comunicação; no processo de transmissão o sinal recebe distorções não intencionais, chamadas de ruído. Chegando no receptor, o sinal é transformado em mensagem novamente.

No caso da telefonia, o canal é um fio, o sinal é a corrente elétrica variante nesse fio; o transmissor é o conjunto de dispositivos (telefone, transmissor etc.) que alteram a pressão do som da voz na variação da corrente elétrica. [...] No discurso oral, a fonte de informação é o cérebro, o transmissor é o mecanismo de voz produzindo a variação de pressão de som (o sinal), que é transmitido pelo ar (o canal).² (SHANNON; WEAVER, 1964, p. 7).

A teoria matemática da comunicação estabelece um modelo de compreensão objetivo do processo comunicacional e apresenta três problemas de pesquisa para essa área do conhecimento: 1) O problema técnico, que procura identificar a exatidão com que os símbolos podem ser transmitidos. 2) O problema semântico, que busca entender como os símbolos transmitidos convergem para o sentido desejado. 3) O problema da efetividade, que investiga como o sentido recebido afeta a conduta do receptor do modo desejado.

A abordagem matemática de compreensão do processo de comunicação é precursora da utilização estatística nos processos automatizados de análise. Shannon (1964) formula a sua teoria, capaz de calcular a probabilidade de aparecimento da letra seguinte dentro de um conjunto de probabilidades, reduzindo a entropia ou o caos esperado. O modelo estatístico de análise é implementado anos depois no processamento de linguagem natural e no *software* livre dadossemiotica.

O ESQUEMA DE COMUNICAÇÃO

A compreensão da comunicação na perspectiva da linguagem recebeu grande contribuição das pesquisas do russo Roman Jakobson, que, em 1958, guiado pelo propósito de entender como o estilo se concretiza na linguagem poética, publicou em uma conferência o manuscrito que estabeleceu seis fatores capazes de classificar a comunicação de acordo com as funções da linguagem, conforme imagem 2.

2. Tradução livre para: In the case of telephony, the channel is a wire, the signal a varying electrical current on this wire; the transmitter is the set of devices (telephone, transmitter, etc.) which change the sound pressure of the voice into the varying electrical current. [...] In oral speech, the information source is the brain, the transmitter is the voice mechanism producing the varying sound pressure (the signal) which is transmitter through the air (the channel). (SHANNON; WEAVER, 1964, p. 7)



*Imagem 2: Esquema de fatores da comunicação verbal
Fonte: Jakobson (1960) adaptado e traduzido pelo autor.*

De acordo com o esquema, a comunicação se dá através de um remetente que envia uma mensagem para um destinatário. Essa mensagem deve ter contexto ou algo a que se refere, um código comum entre remetente e destinatário e um contato, um canal capaz de conectar remetente e destinatário (JAKOBSON, 1960, p. 3).

As funções da linguagem, que aparecem em conteúdos escolares do ensino médio, são parte da iniciativa de Jakobson de compreender como a comunicação verbal se torna arte, resultando na constatação de que a poesia é apenas uma das funções que constituem um evento de fala.

A função referencial, denotativa ou cognitiva se constitui da comunicação orientada ao contexto, narra um acontecimento e admite características descritivas para apresentar aquilo a que se refere. A função emotiva ou expressiva é focada no emissor e na expressão de suas paixões, como felicidade, tristeza, surpresa ou angústia. A comunicação que se volta sobre a mensagem e foca nas questões estilísticas da comunicação verbal constitui a função poética da linguagem. A função conativa é caracterizada de forma predominante pelas expressões imperativas e vocativas; essa função foca em promover a ação do destinatário. Os recursos comunicacionais que servem ao propósito de estabelecer, prolongar ou interromper o contato entre os interlocutores de um discurso caracterizam a função fática da linguagem. Quando o remetente e o destinatário precisam se assegurar de que estão sob o mesmo código, utilizam a função metalinguística (JAKOBSON, 1960).

Essa classificação é estabelecida com base na predominância dos fatores de comunicação (JAKOBSON, 1960 p. 3), o que admite a sobreposição de funções

em uma mesma comunicação. A caracterização deve ser realizada de acordo com a intensidade com que os fatores incidem sobre a mensagem analisada.

O esquema de Jakobson apresenta o entendimento geral do que se espera em uma comunicação e o que precisa ser entendido e implementado pelo remetente para se comunicar com seu destinatário. As funções da linguagem representam em termos gerais a intencionalidade do discurso como parte da semântica do texto e de seu contexto.

O PROCESSO SEMIÓTICO DA COMUNICAÇÃO

A compreensão do processo de comunicação com o auxílio dos modelos e esquemas foi complementada e explicitada ao longo dos anos. Em 2008, Ana Cristina Fricke Matte (2008) se referiu à tese de 1972 de Ignácio Assis Silva, que havia sido citada por Diana Luz Pessoa de Barros (2010), e apresentou a importância de se compreender a comunicação como um processo instável.

O modelo do linguista Ignácio Assis Silva, apresentado e complementado na imagem 3, apresenta três tipos de ruído que levam à compreensão da comunicação como um processo dinâmico. A comunicação, diante dessa compreensão, sempre estará sujeita a “problemas relativos ao uso do código, à manifestação da mensagem em si e ao processo semiótico, especialmente focalizado no conteúdo” (MATTE, 2008, p. 2).

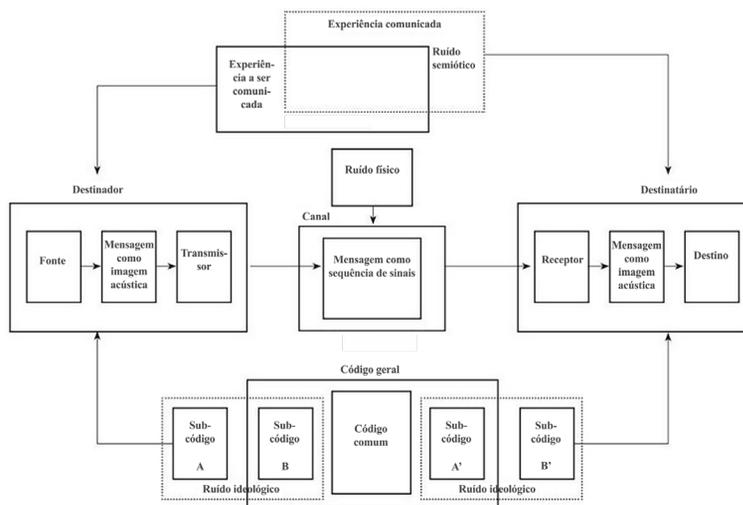


Imagem 3: Modelo de comunicação de Ignácio Assis Silva, com o acréscimo do ruído semiótico
Fonte: Ignácio Assis Silva, citado por Matte (2008) e adaptado pelo autor.

No esquema apresentado na imagem 3, é possível perceber a presença de três eixos, que conectam destinador e destinatário no processo de comunicação: o eixo da experiência no topo, o eixo da mensagem no centro e o eixo do código na parte inferior.

Na explicitação do esquema, Ana Matte (2008) destaca a importância da utilização do termo destinador no lugar de remetente, uma vez que ele pressupõe uma relação em que o destinatário possui um papel ativo no processo. A comunicação bem-sucedida depende de características e competências atribuídas também ao destinatário.

Começando pelo eixo central, ou via da mensagem, o esquema da imagem 3 apresenta complementações mais detalhadas da compreensão matemática de Shannon e Weaver (1964) ou dos fatores de Jakobson (1960). A conversão da mensagem como imagem acústica para uma sequência ou conjunto de sinais (MATTE, 2008) é uma referência à compreensão da comunicação falada de Ferdinand de Saussure (2006), segundo quem os fatos de consciência ou conceitos que um indivíduo possui no cérebro são associados a imagens acústicas e convertidos em signos capazes de serem comunicados. Essas mensagens acústicas seriam, então, expressas pelo sistema fonador ou algum outro transmissor.

O que no esquema da imagem 3 aparece como sendo a “mensagem como sequência de sinais” trata do conceito ou ideia do destinador convertido em um conjunto de sinais. Esses sinais são potencialmente capazes de serem recebidos pelo destinatário; eles são convertidos em “imagens acústicas” novamente e, se tudo tiver corrido bem, compreendidos por esse mesmo destinatário. Na explicação do esquema, a presença do ruído é indissociável do processo de comunicação.

A mensagem, uma vez convertida em conjunto de sinais ao passar pelo meio de contato ou canal, está sujeita, para chegar ao seu destinatário, às interferências do ruído físico, que são distorções, erros ou interferências incidentes no meio. O som de uma serra circular que atravessa a sala de aula dificultando a compreensão do que está sendo dito pelo professor, um erro de digitação, uma tela com pouco contraste entre o texto e o plano de fundo, perda de pacotes de dados na conexão de internet, são exemplos de ruído físico.

A via do código, ou o eixo inferior que conecta destinador e destinatário no esquema de Ignácio Assis Silva, estabelece a existência de um conjunto de códigos comuns entre os interlocutores como condição para a concretização do processo comunicacional. Não é possível haver comunicação entre dois indivíduos que não compartilham o código usado para se comunicar, ou que não possuem um “código comum”.

O código comum aparece como uma entidade acima dos sujeitos, tal como a entidade Língua proposta por Saussure. No entanto, esse código comum aparece no esquema separando o destinador do destinatário. Essa entidade aparentemente fixa, interna mas isolada no processo, funciona como uma referência para os subcódigos, os quais são, na verdade, os grandes atores da relação entre destinador e destinatário na via do código (MATTE, 2008, p. 5).

Os subcódigos do destinador e do destinatário aparecem no esquema como principais responsáveis pelo ruído ideológico que é formado pela inevitável variação dos códigos utilizados no processo de comunicação. O uso de códigos que não são compartilhados entre os interlocutores dificulta a compreensão do que está sendo comunicado.

Variações linguísticas regionais, culturais, gírias, são elementos que compõem o conjunto de variantes particulares que os indivíduos adquirem durante a vida. No caso da comunicação escrita, a experiência com a língua é o fator definidor. Matte (2008) lembra que é possível considerar idade, sexo, região, classe social, como forma de generalizar as variantes.

Os subcódigos aparecem no esquema como elementos definidores dessas variações. Pressupondo-se o interesse na eficiência do processo de comunicação e a busca por minimizar a incidência do ruído ideológico, o destinador e o destinatário se utilizam de subcódigos que acreditam ser mais adequados para o processo.

O destinador que possui o seu “subcódigo A” escolhe deliberadamente desse conjunto apenas o conjunto de códigos que ele acredita compor a variante do seu destinatário, o “subcódigo B”. Do mesmo modo, o destinatário utiliza apenas o conjunto de subcódigos que acredita compor o repertório do destinador, o “subcódigo A”, para interpretar a mensagem.

A inconformidade entre o “subcódigo A” do destinador e o “subcódigo B”, que o destinador acredita ser o do destinatário, e entre o “subcódigo A”, que o destinatário pensa ser do destinador, e o “subcódigo B” do destinatário é uma possibilidade concreta em qualquer processo de comunicação.

Na parte superior da imagem 3 está a via da experiência, ou via semiótica, em que é possível perceber apenas parte da “experiência a ser comunicada” incluída no conjunto da “experiência comunicada”, que também não está totalmente incluída na “experiência a ser comunicada” pelo destinador. A intersecção corresponde apenas a uma parte do que se gostaria de comunicar e daquilo que foi efetivamente comunicado; a diferença entre um e outro, atribuída pelo destinatário, é chamada de ruído semiótico.

Entender o sentido dos textos verbais e não-verbais, objeto de estudo da teoria semiótica, envolve a utilização de técnicas para recuperar informações que não estão expressas de forma concreta no texto. A compreensão do “percurso gerativo de sentido”, metodologia da teoria semiótica, baseia-se na pressuposição lógica do não-dito, através do que é dito.

Um filme com duas horas de duração normalmente é recontado para outra pessoa em 15 minutos; o destinador deixa de contar o que ele acredita que o destinatário é capaz de preencher, sem a sua necessária manifestação (MATTE, 2008, p. 6). Quando esse preenchimento não é feito em perfeita conformidade com o que aconteceu, temos o ruído semiótico.

Esse tipo de ruído se constrói com base nas lacunas da comunicação. A ausência de referência deixada pelo destinador em razão da pressuposição de que o destinatário é capaz de preenchê-las adequadamente constitui uma lacuna, que, uma vez preenchida, pode conter ruído semiótico.

A ação de um indivíduo em uma narrativa é sempre fruto da manipulação do sujeito da ação por algum objeto-valor que não necessariamente aparece no texto: “se um sujeito age, é porque foi persuadido a fazê-lo” (MATTE, 2008, p. 8); esse é um pressuposto da narrativa. Cabe ao destinador estabelecer pressupostos com relações mais próximas do destinatário para que o preenchimento das lacunas seja realizado com menos ruído semiótico: “o preenchimento dessas lacunas pelo destinatário é um requisito para a comunicação” (MATTE, 2008, p. 8).

A comunicação entendida como processo permite que ela seja compreendida dinamicamente, como algo que está constantemente sujeito à ação do ruído. Aos interlocutores resta aceitar a imperfeição e minimizar a sua interferência. Os ruídos semiótico, físico e ideológico são parte do processo e caracterizam a imperfeição tão experienciada por humanos e *softwares*.

DADOS SEMIÓTICA

A iminente e estratégica necessidade de informação e comunicação imposta pelo contexto sociopolítico em que o mundo se organizou durante e depois da Segunda Guerra Mundial foi condicionante para o aprofundamento das pesquisas sobre a linguística computacional, com o objetivo central de buscar resultados para a tradução automática de textos.

A substituição pura e simples das palavras que compõem o texto não representaram um grande problema para os estudos de tradução. O problema a ser superado era a polissemia lexical, que exige do processamento computacional formas de lidar com o contexto em que a palavra é empregada. Diversos modelos foram propostos para contornar essa limitação, como abordagens estatísticas, redes neurais, criptografia e, mais recentemente, técnicas de aprendizagem de máquina. Os estudos para tradução automática permitiram que a linguística computacional se tornasse uma área de conhecimento consolidada, que ainda busca o aprimoramento na tradução, juntamente com os estudos sobre o sentido do texto.

É com o objetivo de análise do sentido do texto que, em 2012, o projeto Texto Livre lança o *software* dadossemiotica, utilizando-se da base teórica da semiótica francesa. O sistema organizado no *software* se fundamenta na cultura livre, caracterizando não só a licença de distribuição, mas também a própria arquitetura do sistema.

A abordagem idealizada pelo Texto Livre é a utilização de um sistema de análises automáticas, com parâmetros estatísticos para segmentar o texto em frases. Através do corretor gramatical (e *software* livre) CoGrOO, utilizado no LibreOffice, o sistema realiza automaticamente a inferência de estruturas como sentenças e palavras, e também o gênero, número, pessoa e tempo dessas últimas. Essa abordagem de processamento da linguagem natural é chamada de tokenização, reduzindo o texto em unidades menores para facilitar a sua consulta e análise.

Mas o desafio contemporâneo do processamento da linguagem natural e da linguística computacional é como lidar com as variáveis para as regras. Não é diferente na programação do *software* dadossemiotica, que, nas situações de ambiguidade, utiliza técnicas estatísticas de aprendizagem de máquina. A abreviação de senhor (“Sr.”), por exemplo, precisa ser entendida pelo *software* como parte da sentença e não como uma nova, marcada pelo ponto final. Na frase “quem casa quer casa”, a ambiguidade morfológica requer um sistema capaz de realizar análise de contexto para distinguir o verbo do substantivo (MATTE; SILVA; CANALLI; RIBEIRO, 2012).

A aplicação dessa tecnologia nos robôs de conversação também faz parte do conjunto de possibilidades do *software* dadossemiotica. Proposto inicialmente como o módulo de processamento de chat, a aplicação foi adaptada para as redes de conversação IRC, com o intuito de analisar a interação dos chatbots de suporte, muito comuns no gerenciamento de salas de *softwares* livres na rede *freenode*.

O *software* está em fase *alpha* e disponível no repositório *sourceforge* (MATTE; MULLER, 2018) para *softwares* livres. A aplicação dos recursos de análise automática

dos textos está sendo direcionada para o ambiente de gerenciamento de aprendizagem Moodle, trazendo mais facilidade e rapidez para o processo de implementação da tecnologia. Com esse tipo de implementação, a tecnologia poderá ser utilizada pelo professor direto na plataforma de aprendizagem, oferecendo ao docente a possibilidade de escolher um texto e compartilhar com os estudantes os resultados da análise do dadossemiotica em diferentes formatos.

RESULTADOS E DISCUSSÕES

Analisar o sentido do texto de forma automatizada e veloz tem relevância para o mundo real e se concretiza em diversas áreas do conhecimento, como educação, marketing e tradução. Para a educação, como é o caso do dadossemiotica, obtém-se na promoção de discussões sobre a compreensão do sentido uma ferramenta que potencializa a análise crítica de textos.

A principal vantagem desse tipo de abordagem analítica é a sua escalabilidade. Um mesmo pesquisador pode analisar um volume de conteúdo muito maior com o auxílio dos recursos de automação. Esse tipo de pesquisa se torna, desse modo, tema de interesse para o marketing digital, que pode se utilizar do processamento de linguagem natural para buscar redes sociais ou áreas de comentários de produtos para melhorar os próprios produtos ou a segmentação dos clientes (GAO *et al.*, 2018).

Além das implementações pragmáticas, o processamento de linguagem natural se aplica em áreas de pesquisa ligadas à análise e à semântica do texto. O estudo da sintaxe do texto e a tokenização, bem como o processo de desambiguação descrito na apresentação do dadossemiotica, compõem esse outro conjunto de possibilidades.

Esse tipo de ferramenta de análise automatizada atua, dentro do esquema de Ignácio Assis Silva mostrado na imagem 3, na “mensagem como sequência de sinais” e pode ter como resultado a análise em nível sintático (tokenização), semântico (classificação das frases) e pragmático (desambiguação) (LENSU, 2002). Além disso, a tecnologia permite implementações capazes de oferecer respostas para o destinatador, como os robôs de conversação, a tradução de textos ou a sugestão de termos nos mecanismos de busca.

CONSIDERAÇÕES FINAIS

Fundamentar-se nos esquemas de comunicação para explicar o dadossemiótica no contexto de estudos do processamento de linguagem natural se mostrou um percurso metodológico relevante para a compreensão dos objetivos do projeto do grupo Texto Livre.

Os problemas que foram abordados pelas iniciativas teóricas e práticas aqui estudadas demonstram a influência do contexto histórico e das áreas de conhecimento dos autores, resultando na interdisciplinaridade das soluções práticas, como o dadossemiótica. O problema de Shannon e Weaver (1964) era o canal de comunicação; Jakobson (1960) buscava compreender as variáveis que compunham a beleza da mensagem; Ignácio Assis Silva (BARROS, 2010) identifica as lacunas e a importância do ruído no processo de comunicação.

As tecnologias de processamento de linguagem natural tiveram grandes avanços com o aprimoramento das tecnologias de aprendizagem de máquina. Redes neurais, abordagens estatísticas ou métodos mistos compõem conjuntos de abordagens que se mantêm centradas em resolver o principal problema: lidar com o imprevisto. Os ruídos e o preenchimento das lacunas da comunicação constituem um enorme desafio para as máquinas, porque também o é para os humanos. Eles são um elemento intrínseco do processo de comunicação.

Compreender esse processo e aceitar sua imperfeição pode ser um caminho para tecnologias mais capazes e competentes em minimizar os problemas provocados pelo ruído. Seja por meio da inversão de papéis de destinatador e destinatário, como fazem os humanos, ou através de treinamentos cada vez mais robustos, como faz o *software* livre dadossemiótica.

REFERÊNCIAS

- BARROS, Diana Luz Pessoa de. A comunicação humana. In: FIORIN, José Luiz. **Introdução à Linguística: I. Objetos Teóricos**. São Paulo: Contexto, 2010. p. 1-227.
- GAO, Yang *et al.* **Alexa, My Love: Analyzing reviews of amazon echo**. Proceedings – 2018 IEEE SmartWorld, Ubiquitous Intelligence and Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People and Smart City Innovations, SmartWorld/UIC/ATC/ScalCom/CBDCom, p. 372–380, 2018.
- JAKOBSON, Roman. Linguistics and poetics. In: T. Sebeok (Ed.). **Style in language**. MA: MIT Press, 1960. p. 350-377.

- LENSU, Anssi. **Computationally intelligent methods for qualitative data analysis**. University of Jyväskylä, 2002.
- MATTE, Ana C. F.; SILVA, W. D. C. M.; CANALLI, H. L.; RIBEIRO, R. T. DadosSemiotica: coleta e processamento de análises semióticas de texto escrito. *In: WORKSHOP SOFTWARE LIVRE*, 2012, Porto Alegre. **Anais...** Porto Alegre: Sociedade Brasileira de Computação, 2012. v. 1. Disponível em: <http://wsl.softwarelivre.org/2012/0010>. Acesso em: 6 nov. 2017.
- MATTE, Ana Cristina Fricke; MULLER, Daniel. **DadosSemiotica**: collector and manager of semiotica annalasis data. 2018. Disponível em: <https://sourceforge.net/projects/dadossemiotica/>. Acesso em: 22 dez. 2021.
- MATTE, Ana Cristina Fricke. O processo semiótico de comunicação: Sobre o Esquema de Comunicação de Ignácio Assis Silva. **CASA: Cadernos de Semiótica Aplicada**, v. 6, n. 2, 2008.
- SAUSSURE, Ferdinand de. **Curso de linguística geral**. 27. ed. São Paulo: Cultrix, 2006.
- SHANNON, Claude E.; WEAVER, Warren. **The Mathematical Theory of Communication**. Urbana: The University Of Illinois Press, 1964.

SOBRE O AUTOR

Paulo Henrique Souto Maior Serrano é professor da Universidade Federal da Paraíba, do curso de Comunicação em Mídias Digitais. Possui graduação em Comunicação Social – Jornalismo pela Universidade Federal da Paraíba (2008), mestrado em Programa de Pós-Graduação em Linguística – UFMG pela Universidade Federal de Minas Gerais (2011) e doutorado em Administração pela Universidade Federal da Paraíba (2019).

E-mail: paulohsms@gmail.com.

ORCID: <https://orcid.org/0000-0002-3713-5190>.

Recebido em 03 de janeiro de 2022 e aprovado em 07 de dezembro de 2022.